orrick

November 22, 2024

Orrick, Herrington & Sutcliffe LLP
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669

+1 415 773 5700
orrick.com

Annette L. Hurst

E ahurst@orrick.com
D +1 415 773 4585
F +1 415 773 5759

Sent via ECF

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl Street
New York, NY 10007-1312

RE: *Authors Guild et al. v. OpenAI, Inc., et al.* (No. 1:23-cv-08292-SHS)
        Alter et al. v. OpenAI Inc., et al. (No.1:23-cv-10211-SHS)

Dear Judge Wang:

Per the Court's request during the October 30th status conference, on behalf of all parties in the above-referenced actions, counsel for Microsoft hereby submits the below updated chart containing a summary of discovery disputes for discussion at the upcoming December 3rd conference.

Given the number of issues included in the chart below, the parties have endeavored to list the issue categories in order of priority, alternating between Plaintiffs and Defendants.

Sincerely,

/s/ *Annette L. Hurst*

Annette L. Hurst

Key: Green represents Plaintiffs' filing; Purple represents Defendants' filing; Orange represents joint filing

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
I.						Disputes Related to OpenAI's Models	
270[1]	11/19/24	TBD	Plaintiffs' Motion to Compel Documents and Training Data Related to Certain OpenAI Models	Class Plaintiffs	OpenAI	Plaintiffs and OpenAI will have each submitted letter briefs by 11/22/24.	**Plaintiffs' position:** Plaintiffs allege that all OpenAI LLMs are trained on Class Works, so Plaintiffs are entitled to discovery on all OpenAI LLMs. OpenAI may not ignore these allegations and limit discovery to a handful of LLMs. OpenAI should be compelled to provide full discovery on its LLMs. **OpenAI's position:** Plaintiffs' case is about specific GPT models that they allege were trained on specific datasets allegedly containing their books. OpenAI is providing extensive discovery into the relevant models, and even into other models not relevant to the claims and not mentioned in the complaint. Unsatisfied, Plaintiffs seek sweeping discovery on models totally untethered to their claims, including purely internal research work. The discovery Plaintiffs seek would be a massive forensic exercise that could take years of work. The Court should deny Plaintiffs' request.
II.						Dispute on Plaintiffs' Collection and Production of Fourth Factor Fair Use Discovery	
264	11/18/24	272	OpenAI's Letter Brief Regarding Sources of Market-Related Discovery	OpenAI	Class Plaintiffs	OpenAI and Plaintiffs have submitted letter briefing.	**OpenAI's position:** Pursuant to Federal Rule of Civil Procedure 34(a)(1), OpenAI seeks an order requiring Plaintiffs to produce highly relevant discovery related to the markets for and value of the works at issue–evidence that is critical to OpenAI's fair use defense–from entities and individuals who possess documents that are within Plaintiffs' control. Specifically, OpenAI seeks an order compelling Plaintiffs to produce such discovery from their literary agents, assistants, publishers, and loan-out companies, all of which possess documents that Plaintiffs have the practical ability to obtain and are thus within Plaintiffs' control. **Plaintiffs' position:** OpenAI's requested relief should be denied for at least three reasons: (1) Plaintiffs have already agreed to produce responsive documents; (2) Plaintiffs collected and have been reviewing documents from numerous business and administrative accounts that the Plaintiffs have access to and do not believe their assistants or loan-out companies have responsive documents that are not already in their possession; and (3) Plaintiffs are willing to make a request to their

---

[1] All docket numbers cited herein are from the *Authors Guild et. al. v. OpenAI, Inc. et. al.* case.

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							agents and publishers for a clearly defined list of documents that Plaintiffs' counsel can negotiate with OpenAI's counsel.
263	11/18/24	274	Plaintiffs Evidence Regarding Markets for the Works-In-Suit	Microsoft	Class Plaintiffs	Microsoft and Plaintiffs have submitted letter briefing.	**Microsoft's position:** Microsoft has served specific, tailored requests on Plaintiffs to assess the market for (including the potential market) and value of the copyrighted works-in-suit (RFP Nos. 10-12). The Individual Plaintiffs have conceded the relevance and agreed to produce documents in their physical possession related to these Requests. The remaining issues are: (1) Author's Guild is refusing to collect and produce all information in its possession; and (2) the Individual Plaintiffs have not agreed to ask those from whom they have a practical ability to obtain the documents—literary agents and publishers—for this information as well. This discovery is critical to assessing the fourth factor of fair use and, by virtue of its representation of authors in the industry, Authors Guild should be compelled to produce all such documents in its possession, and the Individual Plaintiffs should be compelled to seek this discovery from their literary agents and publishers from whom they have the practical ability to obtain information. **Plaintiffs' position:** Apart from the parties more general disagreement over the which evidence is 'critical' for Microsoft's defenses, Microsoft's requested relief should be denied for at least two concrete reasons: (1) the Authors Guild in fact has produced documents related to its "Copyrighted Works" as that term was defined by Microsoft itself in the requests it cited, that is, relating to Mignon Eberhart (the copyrights in whose literary estate is owned by the Guild); and (2) Plaintiffs are willing to make a request to their agents and publishers for a clearly defined list of documents that Plaintiffs' counsel can negotiate with Microsoft's counsel.
232	10/23/24	244	OpenAI's Letter Brief Regarding Harm-Related Interrogatory Response	OpenAI	Class Plaintiffs	OpenAI and Plaintiffs have submitted letter briefing.	**OpenAI's position:** OpenAI seeks an order requiring Plaintiffs to comply with Federal Rule of Civil Procedure 33(d) by either (a) identifying those documents Plaintiffs have produced that reflect harm they claim to have suffered as a result of OpenAI's alleged use of their works to train ChatGPT, or (b) stating that they have produced no such documents. This discovery is critical to, *inter alia*, OpenAI's affirmative fair use defense, which looks at, among other things, any alleged harm to the market for Plaintiffs' works. This is not a contention interrogatory, plaintiffs responded by citing Rule 33(d), and fair use discovery is essential now, not in the future.

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							Plaintiffs' position: As the Court stated, this motion raises a premature damages-related dispute that the Court should not consider at this stage of fact discovery. *See* 10/30/24 Tr. 48:13-23; 79:23-80:11. If the Court considers OpenAI's motion now, the Court should deny it because (1) these are premature contention interrogatories (L.R. 33.3) and (2) regardless of case stage, these interrogatories are overbroad, burdensome, and legally unsupported by courts in this District.
III.			**Disputes Related to Data Working Group**				
267-Sealed 268-Public	11/19/24	TBD-OAI 277-Sealed MSFT[2] 278-Public MSFT	Plaintiffs' Motion to Compel Documents Related to the Data Working Group	Class Plaintiffs	OpenAI & Microsoft	Plaintiffs and Defendants will have submitted letter briefs by 11/22/24.	**Plaintiffs' position:** The Data Working Group is a joint effort by OpenAI and Microsoft to acquire data to train their LLMs. Defendants do not dispute the relevance of these documents related to the Data Working Group but argue that reviewing them would be too burdensome due to privilege issues. Defendants should be compelled to search for responsive documents and to log any privileged documents. **OpenAI's position:** The Data Working Group is a group composed of lawyers and legal professionals, set up to discuss common legal interests. Its documents are privileged. Plaintiffs are wrong that the DWG had a business purpose. Plaintiffs are already receiving non-privileged documents relating to the topics they claim are the relevant topics through non-privileged sources. The effect of Plaintiffs' demand is to require OpenAI to search the files of its lawyers (none of whom are current custodians), with the only result being an extensive privilege log. This is disproportionate to the needs of the case and unduly burdensome. **Microsoft's position:** Microsoft and OpenAI share a common legal interest stemming from litigation over their data acquisition and use. The Data Working Group ("DWG") was formed in support of that common legal interest. The DWG consists of members of Microsoft's and OpenAI's legal teams who jointly discuss legal requirements, legal strategy, and legal developments related to data acquisition and use within the context of their litigation positions. The DWG communications Class Plaintiffs seek are communications in furtherance of this common legal interest and are squarely protected by the common interest privilege. Class Plaintiffs' request to invade this privilege and compel these communications must be denied.

---

[2] Dkt. 276 is Microsoft's Motion to Seal its letter brief and exhibits.

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
IV. Dispute on Plaintiffs' Collection and Production of Evidence of Plaintiffs' Use of Generative AI							
263	11/18/24	274	Evidence Regarding Plaintiffs' Use of Generative AI	Microsoft	Class Plaintiffs	Microsoft and Plaintiffs have submitted letter briefing.	**Microsoft's position:** Plaintiffs have refused to produce documents in response to Microsoft's Request directed at understanding how Plaintiffs themselves are using ChatGPT (RFP No. 23). This Letter Brief is not directed at prompting in connection with Plaintiffs' pre-suit investigation, but rather their general usage of the tool. Evidence of Plaintiffs' use of ChatGPT is directly relevant to Microsoft's fair use and substantial non-infringing uses defenses. Plaintiffs' interactions with ChatGPT (including their prompts) will show the transformative, beneficial, and pro-copyright nature of large language models and their uses—far beyond the cherry-picked uses alleged in Plaintiffs' Consolidated Complaint. It is obvious why providing Microsoft with OpenAI usernames would not resolve the request, since Microsoft has no ability to access the OpenAI system to obtain their user data. **Plaintiffs' position:** Microsoft's requested relief should be denied for at least three reasons: (1) Plaintiffs already agreed to produce documents related to ChatGPT; (2) ChatGPT outputs unrelated to the works at issue are not relevant because Plaintiffs allege that Microsoft enabled OpenAI's largescale copyright infringement in training its LLMs, not that end users of ChatGPT are generating infringing outputs; and (3) Plaintiffs have offered to identify each individual plaintiffs' ChatGPT username, which OpenAI expressly stated would resolve their nearly identical request.
V. Disputes Related to OpenAI Custodians and Custodial Files							
262	11/18/24	275	Targeted Searches of Former Employee Custodial Files (RFP No. 98)	Class Plaintiffs	OpenAI	Plaintiffs and OpenAI have submitted letter briefs.	**Plaintiffs' Position:** Plaintiffs request that OpenAI provide hit counts for a targeted set of search terms—tailored to RFP No. 98—over the custodial files of 14 individuals. The individuals at issue are former high-level OpenAI employees—some of which worked with books as training data—that have raised public concerns relating to the LLMs' harms and risks and the lack of internal controls at OpenAI. The requested relief is narrowly tailored to allow the parties to assess the actual burden of producing documents for these individuals. As the Court noted at the October 30 hearing, running searches and providing hit counts is an appropriate method to streamline discovery disputes based on the assertion of burden. **OpenAI's Position:** Plaintiffs' motion is an improper attempt to circumvent the stipulated ESI Order by asking OpenAI to collect,

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							process, store, and run hit counts across the data of fourteen *non-custodians*. The ESI Order specifies that hit reports are required solely for *custodians*, and Plaintiffs' request is an attempt to get additional custodians in through the back door. In addition, Plaintiffs' request is disproportionate and unnecessary given OpenAI's discovery efforts pertaining to RFP No. 98 and similar requests. OpenAI has already designated 24 custodians (including four individuals *named* in RFP 98), and has agreed to run more than 600 search terms (including the search terms that Plaintiffs requested for RFP 98). OpenAI has searched non-custodial repositories for complaints regarding alleged copyright infringement (the focus of RFP No. 98), and to produce documents responsive to similar topics. There is no reason to believe that these fourteen individuals will have more unique information that is responsive to RFP No. 98, and Plaintiffs have not justified their request to impose this disproportionate and unnecessary burden on OpenAI.
260-Sealed 261-Public	11/18/24	273	Plaintiffs' Motion to Compel Documents Related to OpenAI affiliate defendants	Class Plaintiffs	OpenAI	Plaintiffs and OpenAI have submitted letter briefs.	**Plaintiffs' position:** Plaintiffs have asserted vicarious and contributory infringement claims against several OpenAI entities and seek discovery about these entities' financial ties and control over one another. OpenAI has resisted discovery on the sole basis that, according to OpenAI's attorneys, only one of these entities is relevant. OpenAI says it needs more time to consider what more it will produce, but these requests have been outstanding for four months. **OpenAI's position:** This dispute arose from a failure to meet and confer. The parties agreed to continue discussion about what, if any, additional documents Plaintiffs required for these requests after OpenAI produced, and Plaintiffs had a chance to review, a number of documents that explained OpenAI's corporate structure. Weeks after receiving that production, Plaintiffs made a slew of new demands and, two business days later, filed this motion over OpenAI's objection that the parties needed more time to confer. That time was necessary because all of Plaintiffs' requests are overbroad, seeking whole categories of documents (like documents relating to *any information exchanged* between OpenAI entities) without regard for subject matter or relevance. Plaintiffs' oblique references to needing discovery on secondary liability theories cannot justify the breadth of these requests and the parties should have worked together to arrive at a mutually agreeable scope. Having failed to secure Plaintiffs' cooperation in that effort, OpenAI has done that work itself and set forth, in its opposition,

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							appropriately scoped offers to produce responsive documents that should resolve this dispute.
193 230	9/5/24 10/23/24	199 246	Plaintiffs' Motion to Compel Documents From Additional OpenAI custodians	Class Plaintiffs	OpenAI	Plaintiffs and OpenAI have submitted letter briefs and this issue is subject to forthcoming deadlines.	**Plaintiffs' position:** Plaintiffs seek to add eight OpenAI custodians. These custodians are high-ranking OpenAI officials with relevant knowledge about OpenAI's training data and OpenAI's fair use defense, among other topics. On October 31, 2024, the Court ordered Defendants to "provide hit counts for the Disputed Custodians by filing a status letter no later than November 13, 2024."[3] Dkt. 252 at 2. **OpenAI's position:** OpenAI has compiled and submitted hit counts pursuant to the Court's order. *See* Dkt. 255. OpenAI is in the process of compiling hit counts pursuant to the Court's order. Plaintiffs' request for additional custodians is premature given that they have yet to review the full productions from OpenAI's existing custodians. Plaintiffs also fail to establish that the custodians they request are likely to have relevant, non-duplicative information that would not be available from OpenAI's existing custodians.
VI. Dispute on Request for Early Custodial 30(b)(6) Depositions							
253	11/1/24	253	Plaintiffs' Request For Early 30(b)(6) Depositions of OpenAI and Microsoft Regarding Documents	Plaintiffs	OpenAI & Microsoft	A single, joint, 3-page letter was submitted on this issue.	The Court ordered the parties to confer "regarding the dispute on custodial topics" and file a joint letter by November 1 apprising the Court on whether the dispute has been resolved. Dkt. 252 at 2. The dispute has not been resolved; the parties filed their joint letter at Dkt. 253, and their positions are as follows: **Plaintiffs' position:** Plaintiffs seek two early custodial depositions – one for OpenAI and one for Microsoft. As the Court suggested, an early 30(b)(6) deposition on documents would "help focus document discovery" and resolve many of the parties' disputes. *See* 10/30/24 Tr. 8:2-20. OpenAI has agreed to a custodial deposition with conditions, while Microsoft refuses a custodial 30(b)(6) because "[p]laintiffs have raised no issue with the court relating to Microsoft's custodians, sufficiency of document collection, or document retention." Microsoft is mistaken – Plaintiffs have disputes with Microsoft (including disputes on custodians and the location of documents) that would be advanced by a custodial deposition.

---

[3] At the October 30, 2024 status conference, the Court indicated that these hit counts should reflect "the amount of duplicative documents or the amount of documents for which no other current custodian is a custodian." 10/30/24 Tr. 17:20-18:6.

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							Defendants have Class Plaintiffs' topics, and there is nothing burdensome about a deposition 27 days from now. Because neither Defendant has proposed topics to Plaintiffs, OpenAI's tit-for-tat request that early custodial depositions be reciprocal is inappropriate. For these reasons and those discussed at Dkt. 253, the Court should order both OpenAI and Microsoft to sit for 7-hour custodial depositions by December 3. **OpenAI's position:** OpenAI is conferring with Plaintiffs regarding early custodial 30(b)(6) depositions, but Plaintiffs are seeking a deposition from OpenAI on overbroad, privileged, and improper topics, while refusing to provide any custodial 30(b)(6) themselves. The parties' early custodial 30(b)(6) depositions should be mutual, cabined to seven truly custodial topics per side, limited to three hours each, and take place in January 2025. Consistent with the Court's guidance at the October 30 hearing, OpenAI will serve "a cross 30(b)(6) for plaintiffs" to address the numerous "early issues" with Plaintiffs' document searches and collection "that seem to be holding up some of the discovery." 10/30/24 Tr. at 8:13-15. **Microsoft's position**: There is no basis for a custodial deposition of Microsoft. Plaintiffs have raised no issue with the court relating to Microsoft's custodians, sufficiency of document collection, or document retention, so there is no basis for discovery about discovery from Microsoft. To the extent Plaintiffs seek a deposition relating to output data, that is not a "custodial" deposition; it is a merits deposition. And Plaintiffs did not previously request it; they have no grounds for taking it on short notice before the next conference; and such an accelerated schedule would pose a substantial burden. Additionally, this issue was never fully briefed as Microsoft had ¾ of a page to address its position in the joint letter filed with the Court. Nor did Class Plaintiffs serve any notice, nor seek to meet and confer about any of their proposed deposition topics before the Status Conference. The Court should deny Plaintiffs' requested relief as to Microsoft.
VII.			**Dispute Related to Microsoft's LLMs and Training Data Licensing Materials**				
269-Sealed 271-Public	11/19/24	279	Microsoft's LLMs and material related to Microsoft's	Class Plaintiffs	Microsoft	Plaintiffs and Microsoft will have each submitted	**Plaintiffs' position:** The dispute involves Microsoft's refusal to produce (1) documents showing the datasets that Microsoft used to train its LLMs (RFP No. 41), as well as (2) documents relating to Microsoft's efforts to license LLM training data where a license did not

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
			licensing training data (RFP Nos. 41, 46 and 47)			letter briefs by 11/22/24.	result (RFP Nos. 46 and 47). Contrary to Microsoft's assertion, Microsoft's LLMs are within the scope of Plaintiffs' claims. Indeed, Plaintiffs define their class as those whose works were improperly used to train any of "Defendants' large language models." Compl. ¶¶393-397. Moreover, the requested material is relevant various issues, including the fourth fair use factor, knowledge, damages and willfulness. **Microsoft's position:** Class Plaintiffs improperly seek to greatly expand the scope of a case solely about the training of *OpenAI's LLMs*. Plaintiffs seek discovery about *Microsoft's* LLMs, even though the Complaint is devoid of allegations about them, under the guise of relevance to damages and willfulness to the claims they did make. Plaintiffs are asking the Court to compel Microsoft to disclose (1) documents pertaining to unexecuted data acquisition agreements for recent Microsoft LLMs that bear no relationship to the Open AI LLMs at issue in this case, and (2) documents showing the datasets Microsoft used to train these more recent LLMs. There is no rational relationship between Microsoft's own LLMs, which are not at issue in this case, and any alleged infringement resulting from OpenAI's models. Discovery into these areas would multiply the scope of the case several times over, dragging in models, products, and issues that are irrelevant to the Class Plaintiffs' actual allegations. This request should be denied.
VIII. **Dispute on OpenAI's Production of Texts and X.com Messages**							
230	10/23/24	243	Plaintiffs' Motion to Compel OpenAI To Produce Text Messages and X.com Messages	Class Plaintiffs	OpenAI	Plaintiffs and OpenAI have submitted letter briefs and this issue is subject to forthcoming deadlines.	**Plaintiffs' position:** Plaintiffs moved to compel OpenAI to produce its custodians' text messages and X.com messages. Plaintiffs also requested that OpenAI tell Plaintiffs which of its custodians used text messages and/or social media for work. At the October 30 status conference, the Court ordered OpenAI to "report to Plaintiffs the status of the search on text messages," including "what and whether [Defendants] are producing" by November 6. 10/30/24 Tr. 34:20-23. The Court also permitted Plaintiffs to serve subpoenas on the relevant custodians and interrogatories on both Defendants about the devices and applications their custodians use for work (*see* Tr. 44:22-46:8). Plaintiffs served these interrogatories on November 4. The parties met and conferred on November 6 where (1) OpenAI confirmed that it would produce responsive text messages of Sam Altman and Greg Brockman and (2) OpenAI reiterated its refusal to

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							produce responsive X.com direct messages or identify employees who use X.com direct messages for work. **OpenAI's position:** Plaintiffs' motion seeks text messages residing on personal devices from Sam Altman, Greg Brockman, and Mira Murati (a former employee). OpenAI has requested cooperation from all three individuals and can confirm that it will conduct a reasonable search of Mr. Altman and Mr. Brockman's text messages and produce non-privileged, responsive messages. Ms. Murati is a former employee and has declined to provide OpenAI access to her personal device. OpenAI has confirmed that Ms. Murati's personal counsel will accept service of a proper subpoena, and OpenAI is committed to moving that process forward efficiently. X.com direct messages are not within OpenAI's possession, custody, or control, and the California Labor Code prevents OpenAI from asking employees for access to their personal social media accounts. However, should Plaintiffs serve Rule 45 subpoenas for X.com DMs, OpenAI will cooperate with Plaintiffs and its employees to move that process forward efficiently. OpenAI is in receipt of Plaintiffs' interrogatory, sent via email after the close of business on November 4, and will provide responses and objections consistent with FRCP 33.